Retrospective vs. Concurrent Reports: The Rationale for EMA

Norbert Schwarz

University of Michigan

Self-Reports

- Dominant method for assessing behaviors
- Only method suited for subjective experiences
- Usually retrospective, often covering extensive time periods

"Now, I'd like to read you a short list of different kinds of pain. Please say for each one, on roughly how many days -- if any -- in the last 12 months you have had that type of pain. How many days in the last year have you had headaches?" (NCHS; HIS Supplement.)

The Problem

- Are we asking for things that people can't tell us?
  - Relevant information not accessible in memory
  - Answers based on partial recall, reconstruction, and extensive inferences
- Result:
  - Many systematic biases
  - Generated by a limited number of underlying processes.

Two Solutions

- Better interviewing techniques
  - Some progress made (e.g., Event History Calendars)
  - Opportunities constrained by limits of autobiographical memory.

- Simpler tasks
  - Don’t ask for things people can’t tell you anyway!

Real-Time Data Capture

- Methods assess behavior and experience in real time, close to the event.
  - Record single acts (electronic bottle caps) or extensive concurrent self-reports about behaviors, experiences, and their context (EMA)
- Reduce memory and retrospective judgment problems…
  - … and introduce some new problems.

Report Types

- **Historical Information:** *Ever? First?*
- **Frequency**: *How often?*
- **Intensity**: *How intense, pleasant, painful, etc.?*
- **Change over time:** *More or less…?*
- **Covariation/causation**: *When and why?*

Historical Information

- Examples
  - *Have you ever had an episode of back pain?*
  - *In what year did you first have an episode of back pain?*
  - *How frequently did you fight before you got married?*
- RTDC can not provide this information
- Improved interviewing techniques (e.g., Event History Calendars) can help, within limits.

Frequency

- How often during a specified time period?
- R's strategies depend on the nature of the behavior:
  - Is it ***rare & important*** or ***frequent & mundane***?
  - Is it ***regular*** or ***irregular***?

Frequency: Rare & Important

- Rare and important behaviors can be reported on the basis of autobiographical knowledge…
  - *How often did you get divorced?*
- … or on the basis of a recall & count strategy.
  - *How often did you relocate to another city?*
- RTDC is not suited for such tasks, due to the low frequency of the behavior.

Frequency: Frequent & Mundane

- Frequent behaviors of high similarity blend into generic, knowledge-like representations.
  - *"Having lunch at the cafeteria;" "Seeing my doctor"*

- Such generic summary representations
  - Include rich details about general setting and usual events,
  - but lack time and space markers for specific episodes.
  - Makes "recall & count" impossible.

Frequency: Frequent & Mundane

- Respondents resort to a variety of inference strategies to arrive at a reasonable estimate.

- The choice of strategy depends on
  - Regularity of behavior
  - Context in which the question is presented

Frequency: Frequent, Mundane, & Regular

- When the behavior is **highly regular**, respondents can provide a **rate-based** estimate (Menon, 1994).
  - *Go to church every Sunday. Wash my hair every day…*
- Exceptions get missed.
- By and large, these reports are relatively accurate
  - RTDC is not needed, although often possible

Frequency: Frequent, Mundane, & Irregular

- When the behavior is **irregular**, estimation is the only feasible strategy.
- The resulting reports are highly volatile and depend on the strategy used.
- This is prime territory for RTDC, in particular EMA.

Frequency: Estimation Strategies

- Strategies based on partial recall include
  - Anchoring on earlier report (order effects)
    - *I have headaches more often than heartburn, hence…*
  - Extrapolation from recent incidence
    - *I took pain killers three times today, but this was a bad day. So probably twice a day, times 7 days a week…*
- Results strongly influenced by what comes to mind at the moment.

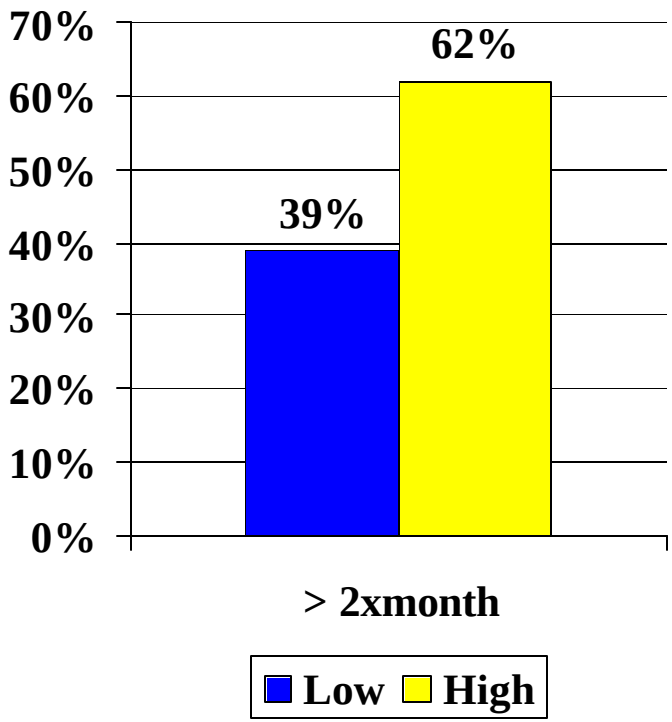
Frequency: Estimation Strategies

- Other strategies largely bypass recall
  - Reliance on information provided by the research instrument
    - E. g., frequency scales
- Throughout, the influence of estimation can be dramatic

Frequency Scales

Low Frequency		High Frequency	
( )	*never*	( )	*twice a month or less*
( )	*about once a year*	( )	*once a week*
( )	*about twice a year*	( )	*twice a week*
( )	*twice a month*	( )	*daily*
( )	*more than twice a month*	( )	*several times a day*

Symptom Reports:
Percent "More Than Twice/Month"



- Patients in psychosomatic clinic
- Averaged over 17 symptoms
- *Schwarz & Scheuring, Z f Kl Ps, 1992*

Frequency: Consequences of Estimation

- Estimation effects increase the more poorly the behavior is represented in memory.

- This undermines comparisons
  - across behavior of differential memorability (e.g., central vs. peripheral symptoms)
  - across groups for whom behavior is differentially relevant
  - across older and younger respondents

Frequency Reports

- **Most behaviors we are interested in are frequent, mundane, & irregular**.
  - For these, retrospective reports are a very poor choice.

- RTDC
  - avoids the memory and estimation problems
  - is highly feasible for frequent events.

Intensity

- Characteristics of subjective experiences, including intensity, are poorly represented in memory.
  - Once the experience ends, it cannot be directly inspected.

- Reports are **constructed** on the basis of
  - limited episodic memory
  - naïve theories about the general type of experience

Intensity

- The experience at the time of report exerts a profound influence on the construction process.

- Direction of influence depends on naive theories of stability or change (Ross, 1989):

  – R's start with present state as benchmark
  – Ask themselves: Was the past similar or different?
  – Adjust their judgment accordingly.

Intensity: "Recency" Effects

- **Stability** (Eich et al., 1985)
  - Chronic pain patients reported current pain and maximum, minimum, usual pain of last week
  - Reports compared to concurrent diary entries

- **Last week's pain more similar to today's pain than warranted**
  - High current pain results in overestimation of past pain
  - Low current pain results in underestimation of past pain
  - But not always…

Intensity: "Improvement" Effects

- **Change** (Linton & Melin, 1982)
  - Back pain patients recorded pain prior to treatment program (baseline measurement)
  - Recalled baseline pain after program completion

- **Retrospective reports show *more* baseline pain than was reported concurrently.**
  - Use present pain as benchmark & adjust based on theory
  - Must have been worse prior to treatment…

Intensity: Stability and Change

- Theory-driven inferences can make the past more or less similar to the present than warranted.

- Particularly problematic when the context suggests the theory: Things get better with treatment!
  - *"You can always get what you want by revising what you had"* (Ross)

- Concurrent measures (RTDC) are the method of choice.

Covariation, Causation, Change

- Self-reports of **covariation** *(Under which circumstances…?)*, **causation** (W*hy…?*) and **change** (*Did it get better?)* pose additional problems.
  - You not only need to monitor behavior (as for frequency judgments) and intensity,
  - but also the variation across time and contexts.

Covariation, Causation, Change

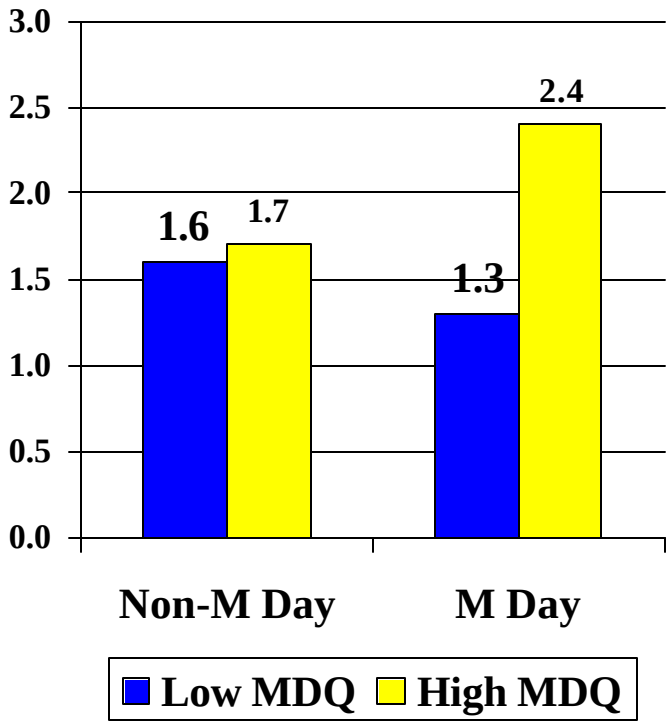
- People are bad at these tasks, even under optimal circumstances.
- R's resort to inference strategies, based on naïve theories of the respective behavior.
  - Numerous systematic biases
  - Can be traced to a small number of underlying processes

Menstruation Beliefs

Example: McFarland et al. (1989)

- Women kept daily diary of affect and physical symptoms
- Later recalled affect and symptoms for a menstruation or non-menstruation day (during intermenstrual phase)
- How do their beliefs about menstruation (assessed with Menstruation Distress Questionnaire) affect the recall?

Diary vs. Recall: Negative Affect



- Shown: Difference Score (Recalled minus Diary NA)
- Higher numbers indicate higher recalled NA relative to diary affect
- Note influence of theory
- *McFarland et a;., 1989*

Covariation, Causation, Change

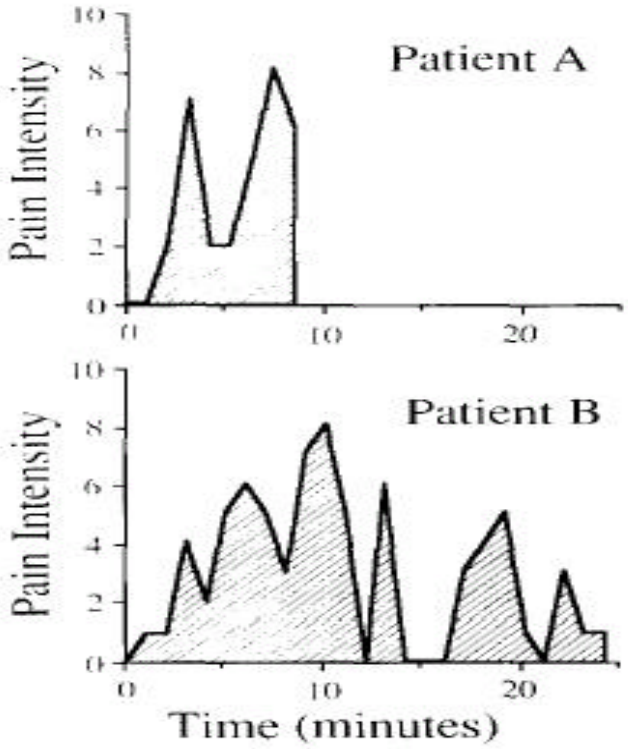
- Reliance naïve theories (beliefs) systematically biases reports of
  - Covariation *(When?)*
  - Causation *(Why?)*
  - Change *(Worse last week?)*
  - Intensity *(How bad?)*

- Except for rare and dramatic events, these reports are not based on episodic recall.

Covariation, Causation, Change

- RTDC avoids these problems by placing the burden where it belongs: on the researcher
  - R's merely report current experiences and behaviors, along with information about the context
  - Assessments of covariation and change, as well as inferences about causation, are based on these data

Evaluating Episodes: Duration Neglect



- Shown: Concurrent ratings of pain during a colonoscopy
- Patient B experiences more pain than patient A
- But in retrospect, Patient B evaluates the episode as ***less*** painful.

- *Redelmeier & Kahneman, 1996*

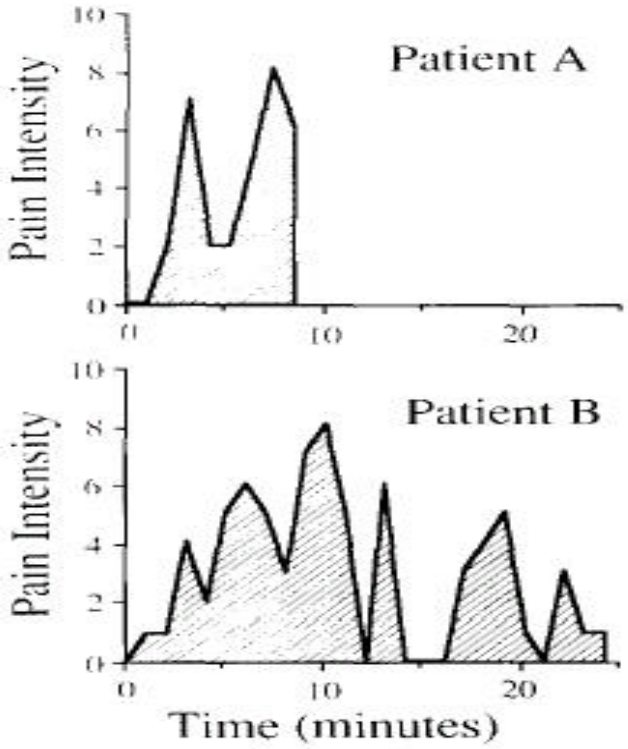
Evaluating Episodes: Duration Neglect

Why?

- Retrospective evaluations follow a **peak & end** heuristic, which draws on 2 pieces of information:
  - How bad does it get? (Peak)
  - How does it end? (End)

- The duration is largely neglected.
  - Judgment not based on "sum" of pain.
  - Report dominated by peak & end.

Evaluating Extended Episodes: Duration Neglect



- Both patients had about the same peak;
- Patient B had a better ending.
- This leaves Patient B with a better memory, despite longer suffering
- … and a higher likelihood to accept a later colonoscopy.

Evaluating Episodes: Duration Neglect

- RTDC can avoid the fallacies of retrospective **peak & end** evaluation
  - But only with dense, concurrent measurement

- Tricky problem:
  - Future behavior is driven by the memory we keep, not by the reality we forget.
  - Does RTDC capture reality, whereas (erroneous) retrospective reports predict behavior in such cases?

RTDC vs. Retrospective Reports

- **RTDC poses a more realistic *cognitive* task and reduces recall and judgment problems.**

- Downsides
  - respondent burden
  - selectivity (respondents & situations)
  - cost

Open Issues

- Biases in retrospective reports are *not* solely due to memory problems and reconstruction:
  - Question interpretation
  - Scale use
  - Social desirability

- We know very little about these problems in the context of RTDC.

Question Interpretation

- Influence of reference period
  - *How often have you been angry yesterday [last month]?*
  - What kind of "anger" is of interest?
  - Less extreme for "yesterday" than "last month"

- Does the short time frame of RTDC invite reports of very minor experiences?
  - Are some of the differences to retrospective reports driven by differences in question interpretation?

Social Desirability

- Negative material is less threatening when it is limited in time and space rather than general
  - *"I couldn't stand my kids last night"* vs. *"I don't like being with my kids."*

- The situation-specific nature of RTDC may decrease social desirability pressure.
  - But for how many repetitions?
  - Do socially desirable responses increase over time?

The Psychology of Concurrent Reports

- Research into the psychology of retrospective reports provides the rationale for RTDC.
  - This rationale is mostly "negative": Avoid the problems of retrospective reports!

- To fully develop the potential of RTDC, we need systematic research into the psychology of concurrent reports.

Some Readings

- Ross, M. (1989). The relation of implicit theories to the construction of personal histories. Psychological Review, 96, 341-357.

- Schwarz, N. (1999). Self-reports: How the questions shape the answers. American Psychologist, 54, 93-105

- Schwarz, N. & Oyserman, D. (2001). Asking questions about behavior: Cognition, communication and questionnaire construction. American Journal of Evaluation, 22, 127-160.

- The latter two are available at:
  http://sitemaker.umich.edu/norbert.schwarz